The BRAHMS data production System

F.Videbæk, H.Ito

July 7, 2004

Abstract

The data production system in BRAHMS has been overhauled to reflect better the ideas move forward during the run-3 D-Au analysis, and in partiucular to have a system that maintains a filecatalog.

1 Introduction

The organization of the production datafiles in run-2, and albeit somewhat improved in the run-3 analysis always lacked clarity an organization. The status of production as well as where to fin the newest official version were murky. During the analysis of run-3 D-Au data, the concept of standardized file location identified by runperiod, specie, and energy proved quite succesfull. During the spring of 2004 there were several discussion between Hiro, Bjoern and myself on how a file-catalog should be organized. The idea is basically that the filecatalog is filled on a per run basis every time datafiles are produced in the ltr (local tracking), gtr (global tracking) or dst (summary tree) step, with the production information, and the location of the generated files. This enables one to ask for any further analysis where the files are located, how they were produced and so on.

Utility cgi scripts have been written so one can query about the prouction version and location of most recent files for any reconstruction step. Methos have been added to the brat-classes in order to add the production files for a given run to a bratmain script.

This document describes the information available in the filecatalog, and how it can be used through usage of standard scripts for reconstruction. as mentioned later it is absolutely essential that these and only these scripts are use for prouction, otherwise we loose the avantage of the filecatalog, and will return to general confusion of placement of isks.

Since RCF is promoting the usage of the Condor batch system over LSF will be phased out due to costs both the crs access as well as the production of gtr and dst has been implemeted using Condor.

The development of the file catalog usage and the utilkity scripts are still under development. At this time though we feel it is ready to be deployed, and can be used for production reconstruction of the run-4 datasets. As of today

(7/7/04) the scripts for dst is not quite done, but the ltr, and gtr scripts for both pp, and auau seems ok. Feedback is of course important to ensure a good working system.

2 Filecatalog Overview

Not yet written.

Information stored. utilities.

3 Guide for Usage of Reconstruction Scripts

There are specific scripts for ltr, gtr and dst production. For a given runperiod they are to be found in `<run><specie><energy>[ltr|gtr|dst]` directories. These are the only official scripts to be used if the integrety and advantage of the filecatalog is to be maintained. please do not use theolder shell scripts except for testing and development. File generated there should be palced in scrath areas.

3.1 Scripts for *ltr*-production

Since all ltr-files are generated by reading raw data stored in HPSS the CRS software developed by RCF must be used. The production scripts uses the most recent CRS software that is deploye on all the farm and uses CONDOR for the underlying batch system.

The official scripts for running ltr production is in the relevant rundirectory e.g. `/run04/auau/200/ltr`.

This text describe how to use masterReductionCondor.rb to submit a job. This script is (hopefully) improvement from the above script in one thing: it automatically update FILECATALOG database. This database includes many information about how and where the files (ltr, gtr, dst, etc...) are created.

This only works at present from rcrsuser3 since it uses Condor..

1 At first, you need to set up enviromental things for Condor by sourcing `setupCrsCondor.sh`. This is usualy done by login time.

2 Then, you can run this script with options. To see the list of options, use `--help` option. A complete listing is in the appendix.

For example, to submit a job for run 11880 - 11882

```
./masterReductionCondor.rb --userName hito --email hito@bnl.gov \
    --beginRunNo 11880 --endRunNo 11882 --productionNo -1 \
    --write-jsf --submit-jsf
```

3.2 scripts for *gtr*-production

These production script submits job to the Condor system. Since the global trak reconstruction the idea of first generating offset files followed by a reconstruction pass w/o having specific files were implemented first with shell scripts by Djamel and for these ruby scripts by Hiro. For a given run the sequences are divided into subsets that are analysis once to generate the offset files, and then to run reconstruction using those files. Thus it takes about twice along, but one is ensure to use a proper set of matchings. (Note we need some script/tools to check the quality ofnthe fit i.e. summary root based scripts).

These scripts are submitted from the cas machines, and will by default run on the CRS nodes.

3.3 Scripts for *dst*-production

example:

For example, to submit a dst generation job for run 11331.

```
./masterDst.rb --userName=videbaek --email=videbaek@bnl.gov \
    --beginRunNo=11331 --inputProductionNo=1 --outputProductionNo= -1
```

The output files i.e. the standard output and standard input can be found in the data directory files associated with the dst output files i.e.

4 Access to filecatalog from brat

The easiest way is use the overloaded `AddFileSet` method to BrIOModule. It uses our FILECATALOG database to find locations of files. You don't need to find and remember the file locations anymore. The format is shown below.

```
AddFileSet(Int_t runno, Char_t* type, Int_t prono, Int_t verno=-1111)
```

The parameters are

runno run number

type "ltr", "gtr", "dst"

prono — production number. The production numbers can be found on `http://pii3.brahms.bnl.gov/daq-cgi-bin/EnterNewProduction.pl`

verno — optional Only if you really want to use the specific version number. Default behavior is to pick the most recent version.

4.0.1 Example.

For example, you want to analyze run 11724 ltr file with production

version number of 1, and will add the following line to the bratmain script

```
inputmodule->AddFileSet(11724, "ltr", 1);
```

5 Script maintanence

For each production all the scripts are really identical apart from the default values. Therefore all scripts (RunInfo.rb,...) are in /run04/scripts/ltr as RunInfo.rb.in etc. These files are commeitted and maintaned in CVS. For each production there is a simple shell script `updateScripts.sh` that takes the scripts from this master directory replaces defaults and puts scripts in `<run><specie><energy>/ltr` ready to use. Thus if changes are to be made in the run by scripts there is ONLY one place now, followed by exceucting the updatescripts for each prouction directory.

TODO: There are several things not in place yet.

1 repeat for gtr scripts (DONE)

1 repeat for dst scripts (DONE)

2 place the common scripts e.g rmdata.rb same place

3 scripts to check consisntency of files on disk - vs filecatalog. This will involve the following tasks. Delete files on official directories if NOT in filecatalog. If files are not on disk

4 scripts to move data to/from HPSS, and to move between different official disk area.

5 Setup so the large local disk on farm machines can be used for dst, gtr files via the `rootd` mechanism.

6 CRS software

The documentation for the CRS software is pretty slim. The following section is taken ad verbatim from the RCF webpage.

See `http://www.rhic.bnl.gov/RCF/UserInfo/Software/CRS-New/`

The New Reconstruction Control Software (New-CRS) handles the scheduling of reconstruction code and defines the environment in which reconstruction code will run. The definition of the environment is outlined in the Reconstruction Control Software Interface document which is available in HTML. The job description in new CRS is identical to the job description in the old system.

6.1 How to use the software.

1. Logon to rcrsuser3 using the reconstruction account.
2. Set the DISPLAY environment variable. Set the xhost permissions on your machine.
3. Execute setup_crs command. This command makes sure that you have the most up to date version of software. You should execute it each time you start working with CRS, at least once per day.

4. Start crs panel: crs_panel [options]. The options define the font type and display width of the panel, you should adjust them to your terminal. Use h option to get help.
5. Prepare a job description file (jdf), exactly like you did it with the old CRS system.
6. Create job using crs_job command, then submit it using submit job command from crs_panel OR create and submit job using crs_job command
7. You can follow the status of jobs using crs panel, or using crs_job -stat command. (The panel is recommended). I will add a web page with job status later.

Description of commands
crs_job [-command] [jdf_file] [-options]
The available commands are:

- -stat : get information about known jobs
- `-create job_description_file`

  $-options$

  Create a job from the job description. Job will be registered in CRS system in CREATED state under a name job_description_file-timestamp. The valid options are:

  `-qn` send job to queue n;

  `-pn` send job with priority n;

  `-drop` allow the job to drop to lower queue if the requested one is full.
- -submit jdf_file : submit a job described in jdf_file. It must be in CREATED state.
- -submit_all : submit all jobs in CREATED status
- -create_and_submit jdf_file: create and start a job.
- -kill jobname : kill job with this name.
- -kill_status status : kill all jobs with given status. For example:
- crs_job kill_status ERROR : kill all jobs in status ERROR
- -reset jobname : reset job, bring it to CREATED state

Examples:

- `crs_job` stat : show list of known CRS jobs

- `crs_job` create jdf_file p10 q2 : create a job from jdf_file, send it to queue 2 with priority 10.
- crs_job create_and_submit jdf_file p10 q2 : create a job from jdf_file, send it to queue 2 with priority 10 and submit the job to condor.
- crs_panel [-options]

  Start crs panel. The options are: -wn set the panel widt to n; -fn use the font n. recommended stettings are -w80 -f10 to get onto most display systems.

7 Appendices

Options

```
"./masterReductionCondor.rb --help"   will produce
You asked to see available options.  See list below
\begin{verbatim}
--userName
--outputDir
--paramFile
--dataDir
--priority
--ffsAngle
--productionNo
--write-jsf
--maxSeq
--mrsAngle
--help
--email
--verbose
--beginRunNo
--configFile
--submit-jsf
--comments
--endRunNo
```

Some options have a default value.

```
configFile = /brahms/u/bramreco/run04/pp/200/ltr/ppLocalTracking.C
paramFile =
outputDir = /brahms/data18/
dataDir   =
mrsAngle  = nil
ffsAngle  = nil
priority  = 10
maxSeq    = 10000
```

The default values are gicen for the pp localtracking. If the defaults are to be changed i.e. outputDir and configFile edit the updateScripts.sh script, execute it and commit to CVS. Do **NOT** edit the ruby script files themselves.

```
Some options are required.
userName
email
beginRunNo
endRunNo
productionNo
```